A common co-ordinate system for mid-sagittal articulatory measurement

James M Scobbie, Eleanor Lawson,
Steve Cowen, Joanne Cleland
and Alan A. Wrench

Working Paper WP-20
June 2011

Queen Margaret University
EDINBURGH

Update and Sourcing Information April 2012

This paper is available online in pdf format

- 2011 onwards at http://www.qmu.ac.uk/casl

Author Contact details:

- 2011 onwards at jscobbie@qmu.ac.uk

Subsequent publication & presentation details:

T none yet

A common co-ordinate system for mid-sagittal articulatory measurement

James M. Scobbie[1], *Eleanor Lawson*[1], *Steve Cowen*[1], *Joanne Cleland*[1] *and Alan A. Wrench*[2]

[1] CASL Research Centre, Queen Margaret University, Edinburgh, Scotland
and [2]Articulate Instruments Ltd.

jscobbie@qmu.ac.uk

Abstract

A standard practice in EMA articulatory measurement is to set the origin of the measurement space near the boundary of the upper incisors and gum, on a standard reference coil. A conventional horizontal dimension is defined as being parallel to the speaker's unique bite (occlusal) plane. We propose that this convention be extended to other instrumentation, with a focus on how it can be achieved for ultrasound tongue imaging (UTI) in particular, using a disposable and hygienic vacuum-formed bite plate of known size. A bite plane trace, like a palate trace, provides a consistent reference to allow images to be rotated and translated in case the probe is in a new location relative to a speaker's cranial space. The bite plane also allows speakers with differently shaped palates to be overlaid, and for ultrasound data to share a coordinate space with EMA. We illustrate the proposal using a sample of six speakers. The average bite plane slope could be used to retrospectively rotate ultrasound data that lacks bite-plane measurement.

Index Terms: articulatory phonetics, EMA, ultrasound tongue imaging, normalization

1. Introduction

There are a number of reasons why in the articulatory phonetic analysis of active articulators in the midsagittal plane it is desirable to relate measurements relative to stable passive articulators, i.e. to fixed locations on the cranium / maxilla. Primarily, it provides a basis for measurements of constriction and the location of the tongue in different segments relative to these locations. A second desirable analytic step is to rotate the image or measurement space into a conventional orientation. This allows short-hand measures in one dimension only, either "horizontal" or "vertical". It is then possible to use phonetic terms like "fronting" or "lowering" in a conventionalized and replicable sense, whether descriptive or quantitative.

Electromagnetic Articulography (EMA) is the most popular technique for articulatory analysis. Previous 2D and contemporary 3D machines capture the location in time, and hence the kinematic properties, of small coils fixed to the articulators. EMA does not, in itself, define a coordinate space for measurement or an orientation of the plane: these are abstract conventions, not physical realities.

A standard aspect of EMA method is to attach a coil to the gum just above the centre of the upper incisors. This area, with relatively immobile skin, acts as a stable anchor-point, and the movement of this coil (and other relatively stable coils on the bridge of the nose and behind the ears) relative to the fixed EMA transmitter electromagnets is subtracted from the other, mobile vocal tract coils, to give the impression of articulator movement within an immobile head. In addition to this necessary function, the location of this coil is conventionally set to be (0,0) in a Cartesian coordinate system intended to measure the midsagittal plane [1, 2].

In addition, at the end of data collection, spare coils (or coils re-used from other locations) are attached to a flat object of indeterminate size suitable for the task, which the speaker in the experiment can bite down on gently. The object is held in the mouth in a unique plane, created by its stable contact against three of the upper teeth, usually one in the left molar/premolar or canine region, one in the right, and the upper incisors. A flat plane comes to rest in a stable unique fashion against the three most pronounced protruding points in any irregularly shaped plane (depending, in the real world, also on the size and weight of the object). In the mouth, at least one lower tooth supports the bite plate when it is bitten against, and holds it steady. This configuration is replicable, because the teeth remain overwhelmingly in the same location on different data-recording sessions (though note, this is not true of longitudinal paediatric contexts). The plane is unique to a speaker, is arbitrary, and is not directly relevant to speech production. However, the occlusal surface of the upper teeth may well be both less variable than the shape of other parts of the face, such as the bridge of the nose or the submental surface, which vary hugely, and being an internal feature, it is likely to correlate with important parts of the vocal tract.

In EMA research, a bite plane (or occlusal plane) is typically captured during data collection. A bite plate might be 4.5 cm wide for an average woman and 5.5 cm for a male vocal tract, of indeterminate length, and made from a variety of materials, even by "laminating a relatively thick paper (at Kinko's, Copy Service, for $1.00-2.00 each)" [3].

As the vocal tract curves through approximately ninety degrees, the concepts of horizontality and verticality, insofar as they make sense, are used primarily in areas of the vocal tract where the natural motions bear some superficial connection to those dimensions, particularly the tongue tip and blade, and the tongue root, but they are pervasive concepts.

In this paper, we show how a bite plane parallel to the EMA-style bite plane can be imaged directly using ultrasound, and propose mechanisms which will allow the imposition of a coordinate space based on this plane, and on indirect measurement of the location of the upper incisor.

Such an approach has two main benefits. The first is that ultrasound data in the midsagittal plane gathered on multiple occasions can be translated and rotated, using the stable bite plane as a guide. This can facilitate series of independent experiments, because it is hard to align data captured on different occasions. One current technique is to rotate and translate based palate traces. The palate is also stable, but is curved, and does not provide either an origin nor a definition of horizontal. Second, ultrasound data (or MRI data, or VICON data, or video-camera data of the face and lips) can be orientated in the same measurement space as EMA data, greatly facilitating our understanding of speech production, given that each articulatory technique has a partial view.

We agree with [2] that "the advantages of standardization could best be served by… uniform implementation of a device" similar to that described here (p106).

2. Bite plate design and manufacture

2.1. Design

A template blank for the bite plate was designed, then manufactured in brass. The requirements were

- A flat under-surface to let the tongue press upwards against the device, visible as a straight line in the ultrasound image.
- Sufficient width to be grasped between the molars, so that the plate records the same plane on every use.
- A transverse barrier on the upper surface to control the distance the plate can be inserted into the mouth, and provide a constant distance from the rear edge.
- An intra-oral gap or protrusion which will deform the tongue away from the bite plane, visible as deviation of the tongue surface away from the bite plane in the ultrasound image in a known location, such as upwards past the rear edge of the bite plate.
- Sufficient rigidity to prevent over-firm upwards pressure from deforming the plate.
- An extra-oral extension of the plate to enable manual grip for insertion and removal of the device, and attachment of EMA coils or motion-capture reflectors, for example, for calibration with other measurement equipment whether or not ultrasound data is being collected.
- Known and replicable thickness and other dimensions.
- Hygienic and safe to use.

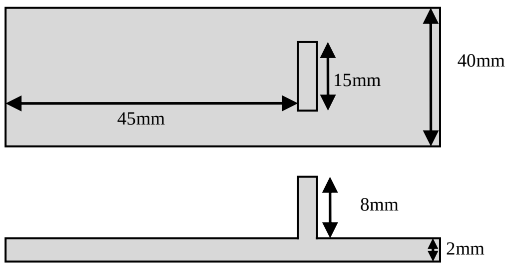

Figure 1: *Schematic diagram of the dimensions and design of the QMU biteplate (version 1). Plan view (top) and side view (bottom).*

This blank can then be used to create individual plastic bite plates (Figure 1) using vacuum-forming. We used 2mm thick HIP (High Impact Polystyrene) dental purpose sheets (10cm by 10cm) which are appropriate for intra-oral use, sterilizable, and cheap. At a cost of approximately 20p per bite plate, they can even be used as disposable consumables if required by a research ethics board. One logistical problem was the time it took to trim the bite plates to size by removing excess plastic, then smoothing the edges, which took 5 to 10 minutes per plate. We estimate the accuracy of the cutting of the rear edge of the bite plate to be within 2mm, as the edges are scored, and cut along a beveled edge. The thickness of the plastic around the upper-incisor transverse barrier is around 1mm, so the brass blank dimensions are smaller than the plastic device shown in Figure 1. In the blank, the upward protrusion is solid, and the plastic deforms over it to make a hollow shape. This helps to strengthen the plate.

3. Experimental trial

3.1. Speakers

Data was acquired from 6 adult female speakers. All participants were native speakers of Scottish-accented or Irish-accented British English. They had never been recorded with ultrasound previously, but all were phonetically trained.

3.2. Ultrasound hardware and data capture

The bite plate can be used with any ultrasound system or with any probe-head stabilization system. Here we describe the set-up for the CASL high-speed UTI laboratory which we used.

Ultrasound data was acquired from an Ultrasonix SonixRP scanner remotely controlled via Ethernet from a PC running Articulate Assistant Advanced™ software [4]. The echo return data was recorded at 100fps with 76 beam-formed echo pulses evenly spread over a 112.5 degree field of view. A hardware pulse was generated by the SonixRP at the instant that each complete set of 76 echo pulses had been recorded. This synchronization pulse sequence was recorded on a multichannel analogue acquisition system at 22,050kHz (along with the acoustic speech signal). The pulses were then detected in a post processing operation allowing each ultrasound frame to be accurately time tagged. A standard graphical interpolation is performed on-the-fly on the stored probe-return data to convert it to an image for analysis within the AAA software environment. This is similar to the image processing that is normally carried out within an ultrasound scanner, but is under the control of AAA. The depth setting was 80mm and the echo return vectors had 412 discrete samples (providing approximately 5 pixels per mm). The transducer frequency was 5MHz providing an axial resolution of approximately 0.9mm.

Recordings were made in a sound-treated studio. Speakers were fitted with a headset (Figure 2) to keep the ultrasound probe within the mid-sagittal plane and stabilize it to reduce movement during speech and between blocks [5].

Figure 2: *Speaker wearing stabilizing headset with ultrasound probe under the chin.*

3.3. Materials and presentation

Single word CVC productions to illustrate the vowel space of each speaker were obtained, as part of other experiments [6]. Real word orthographic materials were presented on cue cards. The speakers were instructed to keep their tongue against their hard palate between tokens, to utter the word in their own time when the prompt appeared, and then to return their tongue to the hard palate. The words contained either /p/ or /b/ as the consonant, and one of eight monophthongal vowels.

The materials were presented in four blocks, in two conditions (upright “U” vs. supine “S”). The condition order was counter-balanced, so the blocks were SUSU or USUS. At the very start and very end of each block, a bite plane reading was taken. The speaker was asked, instead of uttering a word, to insert their bite plate and to press their tongue up against the underside of the plate with the back of the tongue bulging up above the bite plate, if possible, at the posterior edge. Five tokens of *pep* and *pop* were collected in each block, and just one token of the other vowels, in addition to the two bite plane measures.

Here we just analyze upright tokens. We will use the single tracing of the palatal contact from each of the ten tokens of the word *pop* and from the ten tokens of *pep*. The palate tracing tended to be made in the contact preceding the speech token, if possible. We recorded just four upright bite plane measures from each speaker. The headset-probe combination was not moved or removed between blocks, but its passive orientation changed in the supine condition, requiring head correction to be carried out before the upright-supine comparison could be made [6]. However, since it returned to approximately its original position in each speaker's second upright block, we average across blocks without correction.

3.4. Annotation and curve tracing

The following procedures were all carried out within AAA, Version 2.13 [4]. First, a single timepoint was annotated in each token. (Qualitative inspection revealed relatively little apparent movement in the image, and the temporal location of the annotation point was not important.) Second, for each token, a measurement fan grid was superimposed on the fan-shaped ultrasound image, with 42 equally-spaced radial axes. An edge detection algorithm was applied independently along each radial axis to determine the greatest dark-to-light transition, corresponding to the point where the tongue contour crossed the axis. A control point for a spline was added at this point. The algorithm generates a confidence level based on brightness and contrast of the detected edge. Thus each of the 42 axes has one control point, with a confidence level. A spline links these control points on screen, representing the tongue surface. Confidence is indicated visually by fading the tongue contour line where confidence is low. If the user felt the algorithm had made an obvious mistake, the control points were moved manually, and manually given a confidence of 100% (for high confidence of placement) or 0% (equivalent to the control point being removed.

The set of tongue curves for a given experimental condition or speech target can be exported to a spline workspace, where they can be averaged along each radial axis [7]. Thus, along each axis, a mean control point can be determined, and a mean tongue curve plotted to link them. The standard deviation of the sample along the axis is also plotted, and it is widely separated from the mean if data is sparse, variable, or of low confidence. (A threshold of 50% confidence was used here to include data in the averaging process, and the fewer the number of data points relative to the maximum for a mean, the fainter the average curve appears. A curve can even be discontinuous, representing a gap in the raw data, rather than interpolating a curve through empty space.)

For each of the six speakers, one mean bite plane and one mean palate trace was calculated. S1-S4 are reported in [6]. S5 and S6 had more problematic data. S6 only provided 15 tokens for the palate traces, of very poor quality. For both speakers, images of high vowels were weak and lacking clarity, so manual palate tracing was required for these two speakers.

4. Results

The location of the bite plane relative to the probe (in the raw measurement space) is shown in Figure 3. The left point of the bite plane is meaningful, because it corresponds to the back of the bite plate, 45m from the upper incisors along the occlusal plane. The origin and orientation of the coordinate system is, however, arbitrary, based on the size of the field of view in the image window. Different scanners with different settings would locate the tongue in arbitrarily different areas of this measurement space. Perhaps surprisingly, therefore, the individual slopes of the bite plane are relatively consistent.

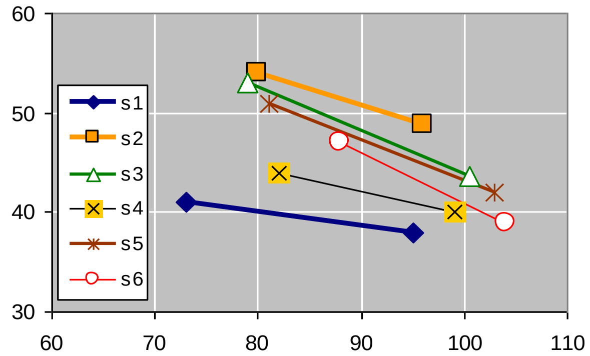

Figure 3: *Bite planes modeled as straight line between two measured points. The left point indicates the back of the bite plate, the right (anterior) point is where the image faded. Origin at bottom left of image area rectangle. Axes in mm. Centre of the probe surface is at (74,10).*

The horizontal axis origin is arbitrary and its origin is set to a point left of (posterior to) the scanner's output image. In this case, the centre of the image is at 74mm, corresponding to the centre of the probe. The vertical axis origin is also set outside the scanner's image, which in this case is at a virtual point at the focus of the fan, 10mm below the tip of the probe. In Figure 3 therefore, the y-axis relates to an origin 10mm below the contact point of the semi-circular probe head with the submental surface. Thus while Figure 3 captures well the varied locations of the bite plane in the measurement space, this variation is purely experimental noise. One reason for the inter-speaker variability is due to the shape and size of each speakers head, and another is the way the probe was placed, including how far it was pushed into the soft submental tissue.

Slightly more usefully, we can calculate the mean vertical distance of the back of the bite plate from the highest point of the surface of the probe itself, and the downward slope forwards of the bite plane. Such measures enable the random variation between speakers and sessions to be quantified.

Table 1 *Variation in the apparent location of the rear point of the bite plane and its slope. Distance is relative to the measurement space*

	s1	s2	s3	s4	s5	s6
Occlusal slope	-8°	-18°	-23°	-13°	-22°	-27°
Distance (mm)	31	44	43	35	42	40
Angular offset	92°	82°	83°	77°	80°	69°

From Table 1, the mean shortest distance from the top centre of the probe to the rear of the bite plane is 39mm (s.d. 5mm) n=6 and the mean angle of this point is an anterior-tilted 81° (where 90° is vertical) with s.d. of 7°.

The random location of the vocal tract within the image is even more clearly displayed if we overlay all the speaker's vowel systems or, as we do here, hard palates. In Figure 4, the location of the hard palate can be seen to vary from speaker-to-speaker, apparently just as much as the bite plane varies.

The dark lines for the palates are means, flanked with thinner lines indicating a ±1 s.d. [7]. The palate traces are different lengths and shapes, which is phonetically meaningful though it also reflects the underlying quality of the image. In Figure 5, the correction has been made by lining up the bite planes, showing more clearly the variation between the measured shapes and locations of the palates.

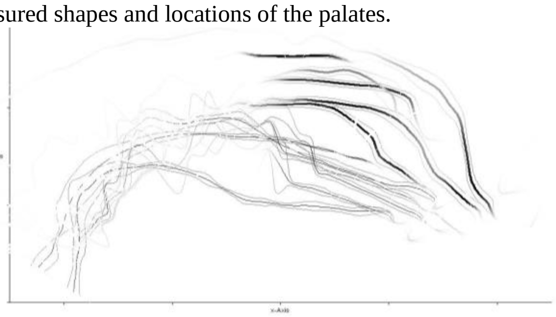

Figure 4: *Six speakers' bite planes and hard palates, uncorrected for probe placement variation*

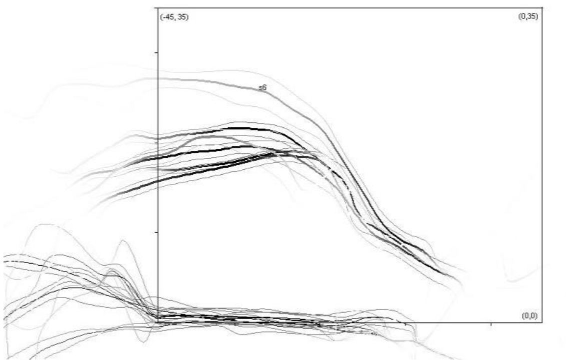

Figure 5: *Six speakers' hard palates, aligned by the bite planes. The bottom right corner is the common origin of the normalized measurement space. The box's horizontal extent is 45ms, the length of the bite plate, and the height, for comparison, is 35mm. Speaker s6's palate is marked. It could be high due to data collection error or physiological difference.*

5. Discussion and Conclusions

The bite plate design could be improved if part of the rear edge could protrude downwards, to give better deformation of the tongue. Some speakers may gag if the plate is too long – a 3.5cm length back from the upper incisor may be preferable to the current 4.5cm. Small ridges could be introduced to the upper surface to increase rigidity. Finally, it may be desirable to reintroduce the more hourglass shape of [2] to aid insertion into the mouth and to be more comfortable when in place, and to have different widths of plate available. Clear extrusions on the external portion of the palate could also be added to help in the placement of EMA coils or motion-capture reflectors, if it is felt desirable for them to be in the same location across sessions. These improvements will be explored in future versions of the bite plate.

The alignment of different speakers' bite planes could be improved also – perhaps by modeling the bite plane as a straight line with two control points (cf Figure 3) rather than fitting a standard curve. It would then be easier to re-orientate those straight lines automatically to a common space.

It will also be necessary to move the vertical origin upwards from the bite plane (which is on the under-surface of the bite plate) if the origin is to be common across UTI and EMA. This requires a speaker-specific measurement of the distance from the plate to the gum above the upper middle incisors, where the EMA origin coil is usually placed. Alternatively, EMA coils could be attached to the underside of a bite plate rather than the upper surface, and the origin moved downwards to that occlusal plane, which would be simpler.

Any stabilization system such as a headset cannot be guaranteed to be re-fitted to the same location on a speakers' head in different trials, but re-orienting images to a common co-ordinate system with a bite plane or palate trace allows for some degree of normalization, useful for matching up longitudinal recordings or for the collection of long datasets across multiple sessions. One minor advantage of a bite plate may be that it is more tolerant to error in the placing of the probe outwith the mid-sagittal plane (though this is an undesirable error, since a different slice of tongue would be being imaged). Small errors in placing the probe in the perfect mid-sagittal orientation will give rise to slightly different palate traces, since the palate is a 3D surface, making superimposition of palate traces more complex. The bite plate, being flat, ought to allow re-alignment even if the probe is translated slightly to a parallel sagittal plane, or has a roll error (but not with a yaw error). It is possible that this property could be exploited to help create 3D images of the palate or tongue movement by compiling and aligning multiple parallel sagittal planes.

Future challenges include correcting dynamic movement of the headset-probe system during speech in a manner which is synchronized with the articulatory and acoustic data streams [5]. This could be achieved by mounting a micro-camera on the headset, and this approach will be pursued in future work, along with refinements to the bite plate design.

[1] Perkell, J.S., Cohen, M.H., Svirsky, M.A., Matthies, M.L., Garabieta, I., and Jackson, M.T.T. Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. Journal of the Acoustical Society of America, 92(6), 3078-3096, 1992.

[2] Alfonso, P.J., Neely, J.R., van Lieshout, P.H.H.M., Hulstijn, W. and Peters, H.F.M. Calibration, Validation, and Hardware-Software Modifications to the Carstens EMMA System. Haskins Laboratories Status Report on Speech Research, SR-114:101-112, 1993.

[3] Cho, T. Electromagnetic Articulography (EMA). UCLA Phonetics Lab Online: http://www.linguistics.ucla.edu/faciliti/facilities/physiology/ema.html, last updated 18 Feb 2000, accessed on 25 Mar 2011.

[4] Articulate Assistant Advanced Ultrasound Module User Manual, Revision 2.12, 2010, Articulate Instruments Ltd.

[5] Scobbie, J.M., Wrench, A.A. and van der Linden, M. Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. Proceedings of ISSP2008, Strasbourg, 373-376, 2008.

[6] Wrench, A.A., Cleland, J. and Scobbie, J.M. An ultrasound protocol for comparing tongue contours: upright vs. supine. Unpublished ms. 2011.

[7] Wrench, A.A., Scobbie, J.M., Cowen, S., Lilienthal, J., Schaeffler, S., Zharkova, N., and Kocjančič, T. Fan-averaging for ultrasound tongue contour evaluation. Oral paper at Ultrafest, Haskins Laboratories, USA., 2010.